

Tests en groupe d'un site web : cas pratique et limites

Nos prestations de test sont souvent comparées aux « Focus groups » et autres études proposées par les principaux instituts ou cabinets de conseil qui font intervenir des utilisateurs en groupe.

En qualité de spécialistes en ergonomie dans l'évaluation de la qualité de l'expérience utilisateur, nous employons une méthodologie adaptée à l'analyse de l'utilisabilité d'un service en ligne.

Nous vous proposons ci-dessous l'analyse d'une étude à laquelle nous avons assisté, analyse qui décrit des problèmes méthodologiques et leurs conséquences sur la qualité de l'évaluation réalisée.

I. Tests en groupe : présentation d'un cas pratique

Nous avons récemment assisté à une matinée de test en groupe organisée par un généraliste du conseil : pendant 3 heures, 10 participants ont évalué la maquette fonctionnelle d'un service en ligne.

Deux groupes de 5 participants ont été conviés successivement aux 2 phases d'évaluation de ce test. Ils avaient au préalable participé aux groupes de travail préliminaires du projet, dont l'objectif était la définition et la validation des fonctionnalités et contenus du service en ligne.

Phase 1 : Les participants étaient placés autour d'une table et devaient effectuer en une heure 5 tâches précises en naviguant dans le site. Les tâches étaient définies selon les principales fonctionnalités du service en ligne. Chaque participant était assisté par un observateur qui l'aidait à comprendre la tâche et évaluait le succès ou l'échec de l'exécution de chacune. Dans cette étude, les observateurs étaient les responsables du projet de service en ligne.

Phase 2 : Il s'agissait d'une séance de travail animée par un consultant. Au cours de cette séance, les participants étaient invités à choisir différents types de graphiques ou de couleurs pour le service en ligne et devaient proposer des solutions permettant de résoudre les difficultés rencontrées lors de l'accomplissement des tâches assignées.

II. Analyse

Si le nombre d'utilisateurs consultés en si peu de temps (et donc à un moindre coût) peut paraître séduisant, la qualité et la pertinence des données recueillies ne nous semble pas permettre d'évaluer réellement la qualité de l'expérience des utilisateurs de ce service en ligne. Les évaluateurs et la méthodologie employée au cours des différentes étapes d'une évaluation sont déterminants de la qualité de l'analyse de l'utilisabilité du service en ligne.

Or, nous avons identifié, dans cette étude, différents problèmes méthodologiques :

- [Non objectivité des participants et du mode de passation](#) :
 - Recrutement des participants
 - Mode de passation des tests
- [Non pertinence des données recueillies](#) :

- Qualification des évaluateurs
- Recueil et analyse des données
- Manque de réalisme des sessions de test :
 - Construction des scénarios
 - Matériel utilisé

Non objectivité des participants et du mode de passation

Recrutement des participants

Les utilisateurs recrutés avaient déjà participé antérieurement aux focus groups au cours desquels les principales fonctionnalités et contenus ont été définis. Ils connaissaient donc le projet et ses principales caractéristiques. Ils ne pouvaient donc pas avoir une vision neutre des informations et services proposés et de l'architecture du site.

La connaissance préalable de l'objectif du service en ligne n'a pas permis d'évaluer sa capacité à répondre aux besoins de ses utilisateurs, plus particulièrement au niveau de la cohérence et de la facilité d'accès aux informations présentées.

Mode de passation des tests

La situation de test est une situation stressante lorsque l'utilisateur ne parvient pas à effectuer la tâche qui lui est demandée. Les participants ont souvent l'impression d'effectuer un test d'aptitude ou se sentent pressés par l'observateur.

Pour le vivre et l'avoir observé au quotidien, nous savons que les utilisateurs ont tendance à culpabiliser et à se dévaloriser. Notre présence à leur côté et notre soutien permettent aux utilisateurs de ne pas se culpabiliser en cas d'échec dans l'accomplissement d'une tâche, en leur expliquant qu'ils ne sont pas responsables de cet échec, mais que c'est le service en ligne qui ne leur permet pas d'accomplir leur tâche.

Ainsi, d'une part, les utilisateurs ne sont pas stressés et démoralisés de l'expérience qu'ils viennent de vivre, d'autre part, nous recueillons des données franches et sincères.

La passation collective génère un stress supplémentaire : la comparaison inter-individuelle : « pourquoi lui y arrive et pas moi ».

Ce type de passation n'est bon ni en terme de données recueillies ni en terme de confort des utilisateurs.

Non pertinence des données recueillies

La qualité de l'analyse de l'utilisabilité d'un service en ligne est très largement liée à la méthodologie de recueil et de traitement des données.

La méthodologie dépend de :

- la qualification des évaluateurs qui l'élaborent et l'emploient pendant l'évaluation du service en ligne
- la pertinence du recueil et de l'analyse des données.

Qualification des évaluateurs

Seuls des experts en Interaction Homme/ Machine peuvent évaluer l'utilisabilité d'un site, c'est-à-dire sa capacité à être facilement compris et utilisable. Il appartient bien à des psychologues cognitivistes formés à l'analyse fonctionnelle des systèmes informatiques plutôt qu'à des spécialistes du marketing et de la sémiologie, de répondre et d'évaluer la qualité de l'expérience de l'utilisateur d'un service en ligne.

Enfin, la compétence en conduite d'évaluation (facilitation du test et entretien) permet de ne pas influencer ou d'induire des réponses spécifiques des utilisateurs.

Analyse

Les observateurs étaient les responsables du projet, et chacun ne voyait qu'un seul utilisateur. Les données recueillies étaient ainsi liées aux différents observateurs, ceux-ci ne s'étant pas coordonnés auparavant. Or, pour avoir un recueil de données homogène et une vision transversale des difficultés rencontrées par les utilisateurs, il est important que tous les participants soient évalués par la(es) même(s) personne(s).

Il en résulte :

- un manque d'homogénéité des données qui pénalise la qualité de l'analyse
- une absence de vision d'ensemble du comportement des utilisateurs.

Recueil et analyse des données

La connaissance de l'interface web et du fonctionnement humain permet de se focaliser sur les données issues de l'observation du comportement et des verbalisations des participants.

La connaissance du fonctionnement cognitif de l'Homme, de ses caractéristiques et limites, permet d'optimiser le recueil des données issues de l'observation de l'interaction entre un utilisateur et un service en ligne, aux niveaux de la pertinence des données recueillies et de la qualité de ce recueil.

Les verbalisations ne permettent pas seules de rendre compte de l'ensemble de l'expérience utilisateur. Par exemple, les informations traitées de manière « automatique », sans s'en rendre compte, sont difficilement verbalisables (e.g. qui est capable de décrire sans oubli et sans erreur les opérations à effectuer pour démarrer une voiture, alors que la plupart d'entre nous le faisons automatiquement ?). De plus, la capacité et les stratégies de mémorisation ne permettent pas de retracer de manière fiable et complète l'ensemble des actions effectuées pour accomplir une tâche.

Quant aux données comportementales recueillies, elles doivent concerner essentiellement les interactions des utilisateurs avec les éléments de l'interface supportant les principales fonctions cognitives.

L'analyse de ces données est ensuite orientée vers l'identification et la compréhension des éléments et événements perturbateurs qui pénalisent l'expérience de l'utilisateur du service en ligne.

Analyse

Les observateurs posaient des questions assez vagues sur des aspects superficiels de l'interface (couleur, style, graphisme) ce qui ne permet pas de recueillir des informations sur les problèmes d'utilisation de l'interface. De même, l'attention des observateurs était portée essentiellement sur les verbalisations de l'utilisateur et non sur ses actes.

Les données recueillies risquent alors de refléter :

- ce que l'utilisateur aimerait entendre ou ce qu'il pense qu'on voudrait entendre de lui
- ce qu'il se souvient avoir fait (mais il ne souvient pas toujours de ce qu'il a fait)
- ce qu'il pense avoir fait et non ce qu'il a réellement fait.

Ainsi, l'identification des difficultés et des stratégies réelles d'utilisation du service en ligne nécessitent la mise en œuvre d'une méthodologie de recueil et d'analyse de données prenant en compte le comportement et les verbalisations des utilisateurs.

La mise en œuvre d'une méthodologie adaptée à l'évaluation de l'expérience utilisateur nécessite la connaissance :

- du fonctionnement habituel d'un utilisateur lors d'une tâche de résolution de problème (par exemple recherche d'information)
- de l'interface web et de la signification et de l'utilisation de ses éléments
- du fonctionnement cognitif humain.

Manque de réalisme des sessions de test

La méthodologie d'évaluation doit permettre de reproduire le comportement le plus réaliste possible des scénarios et du matériel utilisé afin de recueillir des données les plus réalistes possibles.

Construction des scénarios

Le comportement des utilisateurs est différent quand ils ne se sont pas fixé eux-mêmes leur mission : les données recueillies peuvent alors refléter un comportement artificiel parfois très différent du comportement réel. De même, l'ordre de présentation des différents éléments et le temps passé sur les différentes pages du site doivent refléter le plus possible une navigation « naturelle ».

Analyse

Les scénarios et tâches donnés aux participants étaient préalablement définis et ne prenaient pas en compte leurs attentes. Ils sont restés très longtemps sur la page d'accueil, car de nombreuses questions y faisaient référence, alors qu'en réalité, les utilisateurs n'y passent pas plus de 15-20 secondes avant de cliquer sur un lien.

La navigation naturelle doit être privilégiée : le fait qu'un utilisateur ne fasse pas de commentaires ou n'utilise pas un élément du site est un résultat en soi. Cela montre que l'utilisateur ne l'a pas vu ou n'en a pas eu besoin pour accomplir sa tâche.

Matériel utilisé

Le matériel utilisé doit refléter le matériel réel des utilisateurs finals du service en ligne.

Analyse

Les divers ordinateurs mis à la disposition des utilisateurs ne correspondaient pas à l'équipement moyen de la cible visée (ordinateur portable, sans souris, petit clavier, écran réglé en 1024 pixels, connectés au réseau interne de l'entreprise).

Conclusion

L'évaluation d'un service en ligne doit être effectuée par des spécialistes du domaine, et selon une méthodologie rodée et pertinente pour les données à recueillir.

Au final, tout le monde risque d'être pénalisé :

- l'utilisateur final ne sera pas satisfait par le produit réalisé
- le propriétaire du service en ligne ne rentabilisera pas son investissement
- le prestataire du test ne renouvellera pas sa prestation.

Sophie-Laure Pilverdier
Ergonome

Texte retouché et hyperliens ajoutés par le [webmaster](#)